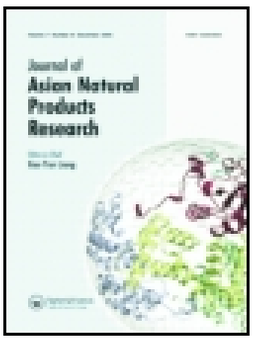


Journal of Asian Natural Products Research

ISSN: 1028-6020 (Print) 1477-2213 (Online) Journal homepage: http://www.tandfonline.com/loi/ganp20

RNA-Seq analysis and comparison of the enzymes involved in ionone synthesis of three cultivars of Osmanthus

Xue-Song Zhang, Jian-Jun Pei, Lin-Guo Zhao, Feng Tang & Xian-Ying Fang

To cite this article: Xue-Song Zhang, Jian-Jun Pei, Lin-Guo Zhao, Feng Tang & Xian-Ying Fang (2018): RNA-Seq analysis and comparison of the enzymes involved in ionone synthesis of three cultivars of Osmanthus, Journal of Asian Natural Products Research, DOI: 10.1080/10286020.2018.1453503

To link to this article: https://doi.org/10.1080/10286020.2018.1453503

JOURNAL OF ASIAN NATURAL PRODUCTS RESEARCH, 2018
https://doi.org/10.1080/10286020.2018.1453503

RNA-Seq analysis and comparison of the enzymes involved in ionone synthesis of three cultivars of *Osmanthus*

Xue-Song Zhang[a,b], Jian-Jun Pei[a], Lin-Guo Zhao[a], Feng Tang[c] and Xian-Ying Fang[a]

[a]Co-Innovation Center for Sustainable Forestry in Southern China, Nanjing Forestry University, Nanjing 210037, China; [b]College of Tea and Food Science and Technology, Jiangsu Polytechnic College of Agriculture and Forestry, Jurong 212400, China; [c]International Centre for Bamboo and Rattan, Beijing 100714, China

ABSTRACT

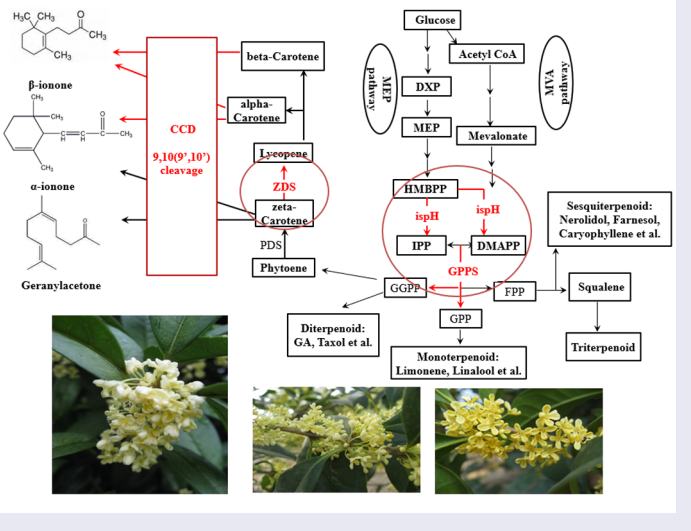
To comprehend the molecular mechanisms that control the differences in the composition of *Osmanthus* essential oils, the RNA-seq data and differentially expressed genes in different cultivar *Osmanthus* were studied. cDNA libraries of "jinqiugui," "baijie," and "rixianggui" were sequenced using Illumina HiSeq ™ 2000. All of the enzymes involved in ionone synthesis were verified. DEGs were revealed and their enriched pathways were analyzed. A total of 20 DEGsencoding four enzymes that were potential candidates involved in ionone biosynthesis, as well as ispH, GPPS, ZDS, and CCD. It provided a way for *Osmanthus* oil monomer material to be synthesized *in vitro*.

ARTICLE HISTORY

Received 12 October 2017
Accepted 14 March 2018

KEYWORDS

Ionone biosynthesis; *Osmanthus fragrans*; terpenoid biosynthesis; transcriptomic analysis



CONTACT Lin-Guo Zhao lgzhao@njfu.edu.cn

The supplementary material for this article is available online at https://doi.org/10.1080/10286020.2018.1453503

1. Introduction

Osmanthus fragrans has been cultivated in China for over 2000 years. It consists of four groups: *O. fragrans* Thunbergii group, *O. fragrans* Latifolius group, *O. fragrans* Semperflorens group, and *O. fragrans* Aurantiacus group. *O. fragrans* essential oil has been paid much attention for its economic and medicinal value. It has also become a new topic of interest for the development of plant resources [1].

The results of many previous studies have indicated that the composition, content and quality of *Osmanthus* oil had the most variation between different origins and varieties. β-Ionone is an important component of the aroma of *Osmanthus*, and its content varies between the distinct varieties of *Osmanthus*. In the Thunbergii group, Latifolius group, and Semperflorens group, the concentration of β-ionone is generally higher than that in the Aurantiacus group [2]. The content of the β-ionone "rixianggui" is the highest in the Semperflorens group, and there is a significant difference in the oil contents of "jinqiugui" and "baijie" [3]. However, there has been little research on the differences in the oil composition and metabolic engineering of volatiles in *Osmanthus* oils on a transcriptomic level.

Previous studies have shown that most floral scent compounds belong to three major groups, phenylpropanoids, fatty acid derivatives, and terpenoids. The commercial and ecological importance of terpenoids makes their metabolic engineering the most valuable of the three groups. β-Ionone, which is the main aroma substance in *Osmanthus* essential oil, is derived from the enzymatic cleavage of carotenoids catalyzed by members of the carotenoid cleavage dioxygenase (CCD) family [4]. In plants, the carotenoid-biosynthetic pathway (Figure 1) has been established [5].

It has become common to use RNA-seq to quantitatively study differences in gene expression in non-model species. By transcriptome sequencing, Mu [6] found 434 candidate genes

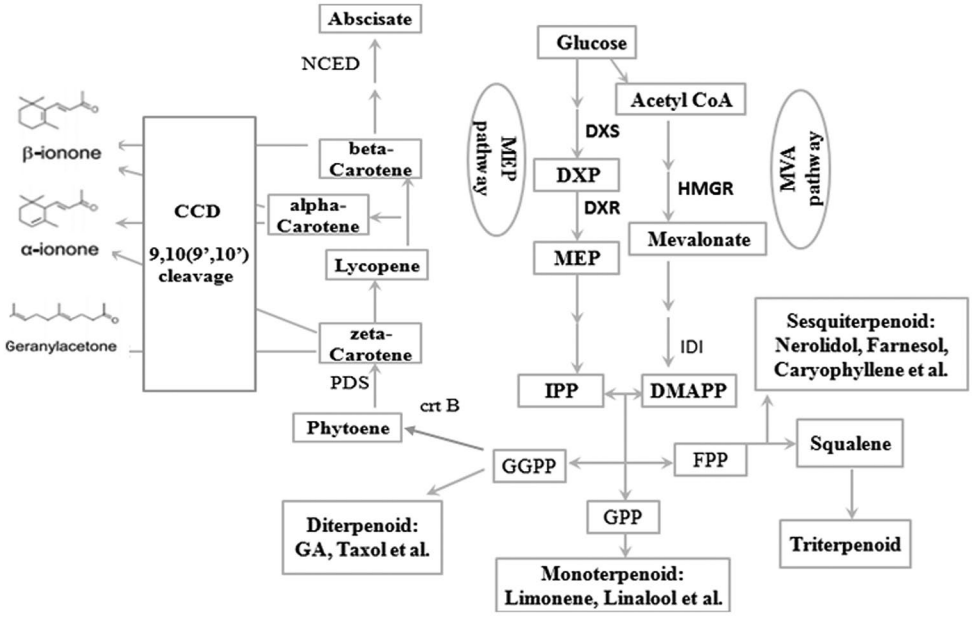


Figure 1. Carotenoid biosynthetic pathway in higher plants.

involved in fragrance and pigment biosynthesis between "zaoyingui" and "chenghongdangui." "Jinqiugui" is a particularly excellent and valuable variety of the Thunbergii group, "baijie" is a high-quality fragrance of the Latifolius group, and "rixianggui" is a valuable, strongly aromatic variety of the Semperflorens group.

This study used RNA-seq to analyze differentially expressed genes in order to better comprehend the molecular mechanisms controlling the differences between their essential oil compositions. Three representative *Osmanthus* varieties, "jinqiugui," "baijie," and "rixianggui," were selected for use in this study. Samples were collected while *Osmanthus* were in the early flowering phase. In the early phase, the main aroma components, such as linalool, γ-decalactone, and β-ionone, have the highest contents [7].

2. Results and discussion

2.1. Component analysis in three cultivars of O. fragrans flowers

To identify differentially expressed genes between "jinqiugui," "baijie," and "rixianggui," the differences in the relative contents of β-ionone in the fragrance compositions of three cultivars were defined first (Supplementary file 1 and Figure 2).

2.2. Illumina sequencing and de novo assembly

cDNA libraries were generated from RNA isolated from "jinqiugui," "baijie," and "rixianggui." Then, paired-end sequencing was performed using the Illumina platform (SRA accession number SRP057917). Each library produced approximately 4.8 gigabases (Gb) of raw data (Figure 3 and Supplementary file 2). In the results of the assembly, 70,029 unigenes were detected. Their average length was 762 nt, and the N50 value was 1183 nt. The bases for "jinqiugui," "baijie," and "rixianggui" contained 53,789,964, 52,898,396 and

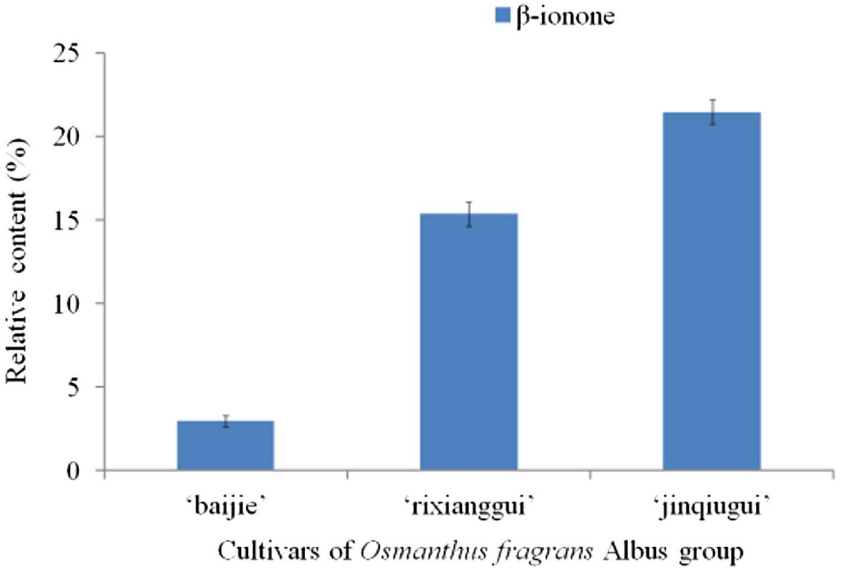


Figure 2. Content of β-ionone of three cultivars of *Osmanthus fragrans*.

Samples	jinqiugui	baijie	rixianggui
			
Total Clean Reads	53,789,964	52,898,396	53,448,460
Total Clean Nucleotides (nt)	4,841,096,760	4,760,855,640	4,810,361,400
Q20 percentage	98.03%	98.04%	98.07%
Number of Unigenes	69,342	74,818	54,735
Mean Length (nt)	588	601	494
N50	960	991	732

Figure 3. Sequencing output of *O. fragrans* cDNA library.

53,448,460 clean reads, respectively. After assembly, 129,896, 143,257, and 105,169 contigs with respective mean lengths of 306, 305, and 285 nt were identified from the "jinqiugui," "baijie," and "rixianggui" samples. After clustering was performed with Trinity software, these contigs were further assembled into 69,342, 74,818, and 54,735 unigenes with mean lengths of 588, 601, and 494 nt, respectively, including 25,839, 28,625, and 18,743 clusters and 43,503, 46,193, and 35,992 singletons, respectively. With a Q20 percentage over 98%, the N50 values of unigenes for the three samples were 960, 991, and 732, respectively. These numbers all indicated that the sequencing data were of high quality to proceed with further analysis. In addition, gene coverage can reflect the quality of sequencing. The genes that had more than 80% coverage by clean reads accounted for 87, 91, and 69% of the total genes of "jinqiugui," "baijie," and "rixianggui," respectively.

2.3. Annotation of predicted proteins

In the assembly, 70,029 unigenes were detected. For functional annotation analysis, 50,980, 42,807, and 32,285 unigenes, respectively, were obtained. The total number of annotated unigenes was 52,869. A total of 67% of the unigenes received high-level hits (E-value < 10^{-30}). The transcripts encoding enzymes of the TCA cycle could be verified by BLAST queries against assemble unigenes (Table 1).

Table 1. *Osmanthus fragrans* tricarboxylic acid cycle (TCA) encoding transcripts.

TCA cycle enzyme	Arabidopsis query	Contig (identity)	Reads
Citrate synthase	AT2G42790.1	CL8688.Contig1_All (80%)	8073
Aconitase	AT2G05710.1	CL305.Contig1_All(80%)	23,009
Isocitrate dehydrogenase	AT1G65930.1	CL4224.Contig5_All (81%)	30,772
2-oxoglutarate dehydrogenase	AT3G55410.1	CL9742.Contig1_All (82%)	15,880
Dihydrolipoyllysine succinyltransferase	AT5G55070.1	CL2198.Contig3_All (80%)	5031
Dihydrolipoyl dehydrogenase	AT4G24160.1	CL2332.Contig3_All (80%)	6154
Succinyl-CoA synthetase	AT2G20420.1	Ungene20456_All(80%)	4487
Succinate dehydrogenase flavoprotein subunit	AT5G66760.1	CL6409.Contig2_All (81%)	9739
Succinate dehydrogenase iron-sulfur subnunit	AT3G27380.2	Unigene13280_All(81%)	8488
Fumarase	AT2G47510.2	Unigene3871_All(80%)	4039
Malate dehydrogenase	AT1G04410.1	CL5475.Contig2_All (80%)	34,837

2.4. GO and KEGG analysis of RNA-Seq data

Based on the NR annotation, 39,082 unigenes could be annotated using GO. They could then be sorted into three ontologies: molecular function, cellular component, and biological process (Figure 4). In each of the three main categories of GO classification, "metabolic process," "organelle," "cell," "cellular process," and "cell part" terms were dominant. Categories of "biological regulation," "catalytic activity," "response to stimulus," "single-organism process," "membrane," and "binding" also showed high percentages. The terms were represented by a few genes, such as "metallochaperone activity," "translation regulator activity," "channel regulator activity," "virion part," "extracellular region part," "extracellular matrix part," "proteintag," "virion," and "extracellular matrix."

A total of 28,057 unigenes were mapped into 128 KEGG pathways (Supplementary file 3). The metabolic pathways had the highest unigene representation (Ko01100, 5874 unigenes, 20.94%), and the biosynthesis of secondary metabolites (Ko01110, 2956 unigenes, 10.54%) and plant-pathogen interaction (Ko04626, 1652 unigenes, 5.89%) are not far behind. The pathways with the least representation by unigene sequences were betalain biosynthesis (Ko00965, 3 unigenes, 0.01%) and caffeine metabolism (Ko00232, 3 unigenes, 0.01%). These annotations will be valuable to subsequent investigation of specific processes, functionsm and pathways in *O. fragrans.* Based on the significant enrichment of particular pathways, the main biochemical pathways and signal transduction pathways that the DEGs took part in were mapped. In the jinqiugui-vs-baijie, jinqiugui-vs-rixianggui and rixianggui-vs-baijie comparisons, 6259, 7770, and 9981 DEGs mapped to 127, 127, and 128 pathways, respectively (Supplementary file 4). Of the 6259 DEGs in the jinqiugui-vs-baijie comparison, 4987 (79.7%) were significantly enriched in 30 pathways. When the jinqiugui and rixianggui libraries were compared, 5669 (72.8%) DEGs were identified in 24 pathways, and in the rixianggui-vs-baijie comparison, 4808 (48.2%) DEGs were significantly enriched in nine pathways.

2.5. Analysis of DEGs in ionone biosynthesis based on RNA-Seq

The transcriptome provided included each of the genes encoding enzymes in ionone biosynthesis (Supplementary file 5). All of these unigenes are important resources for future *O. fragrans* genetic engineering work regarding the biosynthesis of terpenoids. As shown in Supplementary file 5, the 24 enzyme genes involved in ionone biosynthesis had very small differences in expression levels of mvaK2, ispD, and Z-ISO among the three samples. Between "jinqiugui" and "baijie," the MVD, DXS, ispE, and ispF genes also revealed very low differential expression levels. There were significant differences in the expression of the remaining 17 genes between each pair of samples. AACT, MVK, and lcyB were down-regulated, while ispG, IDI, GGPS, crtB, PDS, cruA, and crtISO were up-regulated. For the other seven genes, some unigenes were up-regulated and some unigenes were down-regulated. For example, HMGR is the key enzyme in the mevalonate pathway of terpenoid biosynthesis. Only four unigenes displayed a difference; three were up-regulated (Unigene30816_All, Unigene30345_All, and CL10210.Contig3_All), and one was down-regulated (Unigene26440_All). Additionally, in both the rixianggui-vs-baijie and jinqiugui-vs-rixianggui comparisons, ACAT, ispG, and PDS had very low differential expression levels. MVK, ispF, and lcyB were down-regulated and ispE, ispH, crtB, and crtISO were

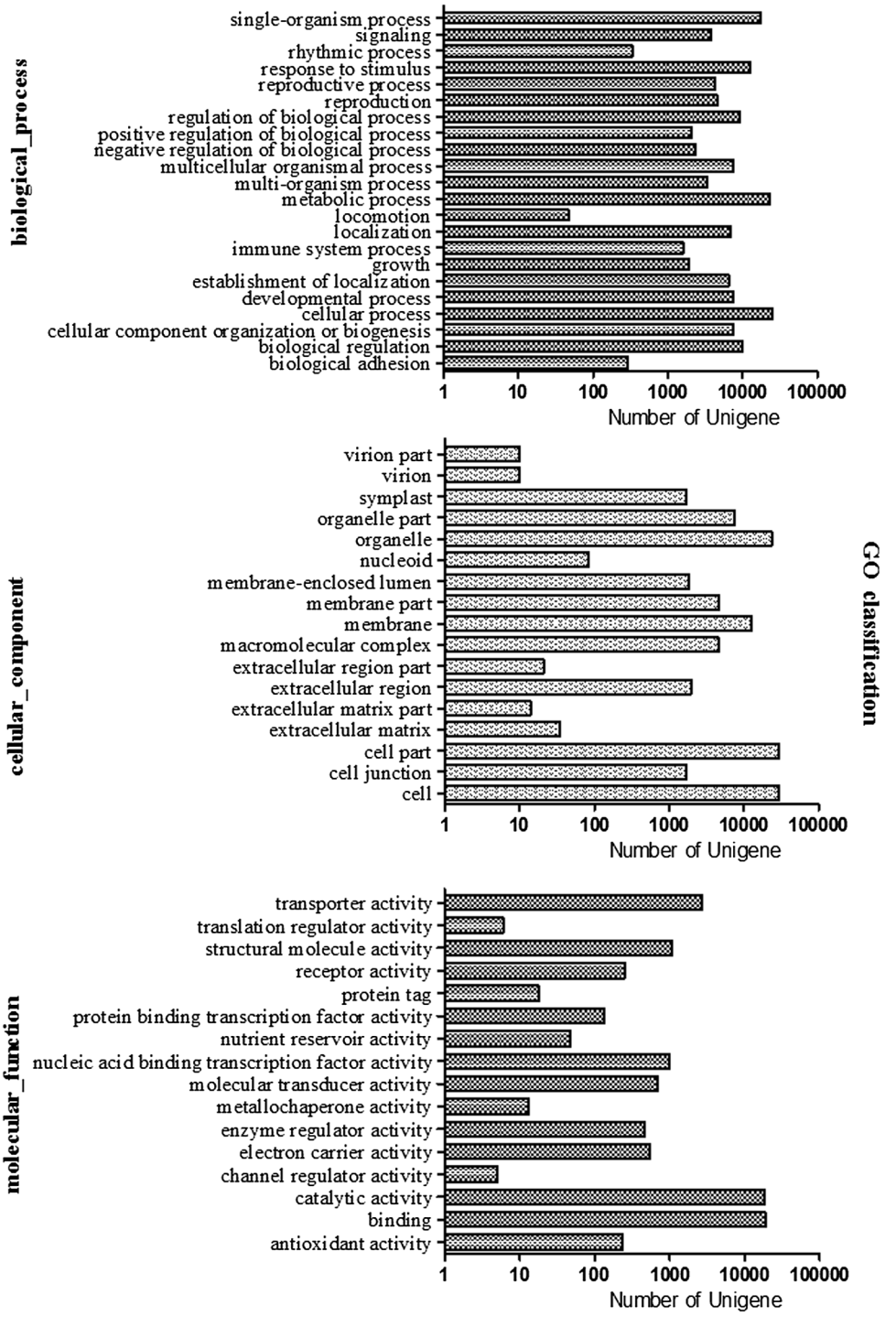


Figure 4. GO classification analysis of unigenes in All-Unigene.

up-regulated in "rixianggui" compared to "baijie." In "jinqiugui" compared to "rixianggui," HMGR, ispF, ispE, IDI, crtISO, and cruA were up-regulated, while MVK, crtB, and lcyB were down-regulated.

The syntheses of various volatile compounds of flowers are all catalyzed by related enzymes, and almost all study of the key enzymes and the genes encoding these enzymes has been done through model plants, such as *Clarkia breweri*, *Antirrhinum majus*, *Rosa* ssp. and *Petunia* hybrids. In recent years, with the development of transcriptomics, studies on the metabolism of aromatic substances have extended beyond model plants. Here, we indentified all of the enzymes involved in the synthesis of β-ionone, a major aromatic substance in *Osmanthus*, and we identified differentially expressed genes at the transcription level along with their enriched pathways in three kinds of *Osmanthus* transcriptomes. The differential expression of enzyme genes may be the cause of concentration differences of β-ionone between *Osmanthus* varieties. According to a previous report [8], transcription levels of carotenoid synthetic genes in the petals of *Tagetes erecta* determined the concentration range of each variety.

2.6. Candidate genes for ionone biosynthesis

In each comparison, 20 unigenes that encoded 4 enzymes had the same or inverse trend in terms of ionone content (Table 2). For precursors of the terpenoid biosynthetic pathway, three candidates were chosen. CL4073.Contig1_All and CL1997.Contig1_All unigenes, which encode ispH and GPPS, exhibited an inverse trend in terms of their β-ionone content. CL1997.Contig5_All, which encoded GPPS, had the same trend. Sequence alignment showed that the sequence of CL4073.Contig1_All was identical to the partial sequences of hydroxymethylbutenyl diphosphate reductase of *Camptotheca acuminate* (GenBank accession number ABI64152.1) and *Camellia sinensis* (GenBank accession number AFQ98369.1). CL1997.Contig1_All and CL1997.Contig5_All were identical to the partial sequences of the geranyl diphosphate synthase large subunit of *Catharanthus roseus* (GenBank accession

Table 2. Candidate unigenes of ionone biosynthesis.

		rixianggui-vs-baijie
		jinqiugui-vs-rixianggui
Gene	Unigene change	jinqiugui-vs-baijie
4-hydroxy-3-methylbut-2-enyl diphosphate reductase (ispH)	Up-regulated unigenes	CL4073.Contig1_All
Geranyl diphosphate synthase (GPPS)	Up-regulated unigenes	CL1997.Contig1_All
	Down-regulated unigenes	CL1997.Contig5_All
Zeta-carotene desaturase (ZDS)	Down-regulated unigenes	CL10334.Contig1_All
Carotenoid cleavage dioxygenase (CCD)	Up-regulated unigenes	CL11071.Contig1_All; CL1680.Contig1_All CL1680.Contig2_All; CL471.Contig2_All CL471.Contig3_All; CL473.Contig6_All CL7905.Contig2_All; CL8956.Contig1_All CL8956.Contig2_All; Unigene4631_All Unigene23542_All; Unigene23625_All Unigene23620_All
	Down-regulated unigenes	Unigene9443_All; CL9444.Contig2_All CL10154.Contig2_All

number AGL91645.1) and ggpps protein of *Jasminum sambac* (GenBank accession number AIY24421.1).

In plant cells, the MVA pathway and the MEP pathway are found in the cytosol and chloroplast, respectively, and are believed to have different physiological functions. The MVA pathway produces only IPP, but the MEP pathway produces IPP and DMAPP. Different ratios of IPP/DMAPP are required for the specific isoprenyl diphosphate biosynthesis in the cytosol and chloroplast, and apparently more IPP than DMAPP is required for the synthesis of geranyl diphosphate (GPP), farnesyl diphosphate (FPP), and other compounds [9]. Hence, maintaining the optimum metabolic flux of IPP/DMAPP in the cell is critical. IspH catalyzed the reductive dehydroxylation of HMBPP to isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) during the final step of the MEP pathway. Our sequence data for *Osmanthus fragrans* show that ispH is a key enzyme in ionone synthesis. Through the regulation of expression of the ispH gene which encodes an enzyme in the MEP pathway, Zhao [10] discovered that under appropriate intensity control of elements M1-64, the production of β-carotene increased by 20%. In addition, GPPS catalyzes a molecular IPP and a molecular DMAPP to synthesize GPP, the precursor of monoterpene. However, Wang's research [11] shows that GPPS strongly affects the GGPS activity of catalytic synthesis of the diterpenoid precursor GGPP by forming a hetero dimmer with GGPS. Furthermore, our transcriptome data show that GPPS is encoded by two different genes simultaneously, and these genes were regulated oppositely. This shows that the GPPS enzymes of *Osmanthus* may have double function identical to the GPPS of orchid [12].

CL10334.Contig1_All, which encodes ZDS, exhibited the same trend as the β-ionone content in each comparison. The sequence of this gene was identical to the partial sequences of CDP13237.1 and EYU30623.1. Isaacson [13] regarded ZDS as responsible for a vital enzyme of the desaturation effect in plant carotene synthesis, and it takes part in linear carotenoid biosynthesis. By cloning the whole cDNA sequence of PDS and ZDS in ripe citrus fruit, Rodrigo [14] discovered that ZDS positively regulates the accumulation of carotenoids in citrus fruits. Zhu's research [15] showed that overexpression of ZDS raised the content of carotenoids in the petals of yellow flowers.

A total of 171 unigenes were annotated as CCDs in the *O. fragrans* transcriptome and DEGs were found in three comparisons (Supplementary file 6). These were all identified as either CCD1 or CCD4. CCD proteins comprise the largest family of plant proteins. In the biosynthetic pathway of ionone, the 9, 10 (9′, 10′) double bond of β-carotene can be cleaved to form β-ionone. CCD8, CCD7, CCD4, and CCD1 from *Arabidopsis thaliana* and *Rosa damascene,* among others, have been identified, and they cleave the 9, 10 (9′, 10′) double bond of carotenoids [16]. According to Table 2, 16 DEGs out of 171 unigenes had the same trend as the β-ionone content or the inverse trend in both the rixianggui-vs-baijie comparison and the jinqiugui-vs-rixianggui comparison as well as in the jinqiugui-vs-baijie comparison. Among these genes, CL11071.Contig1_All, CL1680.Contig1_All, CL1680.Contig2_All, CL471.Contig2_All, CL471.Contig3_All, CL473.Contig6_All, CL7905.Contig2_All, CL8956.Contig1_All, CL8956.Contig2_All, Unigene4631_All, Unigene23542_All, Unigene23625_All, and Unigene23620_All were all up-regulated in each comparison. These exhibited an inverse trend relative to the β-ionone content. Unigene9443_All, CL9444.Contig2_All, and CL10154.Contig2_All were down-regulated and had the same trend as the β-ionone content. Their sequences were also similar to the partial sequence of CCD4 of *Osmanthus fragrans* (GenBank accession number ABY60887.1) and *Rosa x mascene*

(GenBank accession number ABY60886.1). CCD4 enzymes can only cleave apocarotenoids and cyclic none-polar carotenoids, such as β-carotene [4]. In recent research, no studies have investigated the enzymatic carotenoid cleavage included in the determination of the relative levels of ionone [16]. The carotenoid cleavage oxygenases CCD1 and CCD4 have been reported to exhibit high expression levels in flowers and to be involved in the biosynthesis of terpenes, such as β-ionone [17]. Similarly, our sequence data of *Osmanthus fragrans* shows different expression of the CCD4 transcripts in three *Osmanthus fragrans* flowers.

Spatial and temporal disparities in volatile compounds of same species could exist in different parts of a flower organ, different flowers of the same plant or different plants and groups [18]. Due to complex spatio-temporal modulation, the composition and content of volatile compounds released by plants in different flowering periods can change [19]. According to previously reported data, three cultivars were chosen in the early flowering phase for high β-ionone content, and the DEG transcripts accumulated to high levels in the flower, more specifically, during early flower opening [20]. In addition, most of the genes encoding enzymes associated with aroma are expressed abundantly in the petals. A few enzyme-encoding genes expressed in the nutritive organ, such as PhCCD1 gene in *Petunia* [21]. Actually, as a volatile compound, β-ionone is synthesized under developmental and spatio-temporal regulation [22], and this synthesis relies on substrate content in the final reaction procedure. The quantity and efficiency of a substrate is particularly important for the synthesis of volatile substances when the final reaction is catalyzed by an enzyme with broad substrate specificity [23]. When different enzymes compete for the same finite substrate, the composition of the final product depends on the km value of each enzyme acting on the substrate. The synthesis of volatile substances by plants is also controlled by the circadian clock [24], suggesting that the synthesis of β-ionone is extremely complex and requires further study.

Five genes of CCDs were selected to confirm the RNA-Seq results by real-time quantitative PCR (Figure 5). The results were the same as the sequencing data of the three samples.

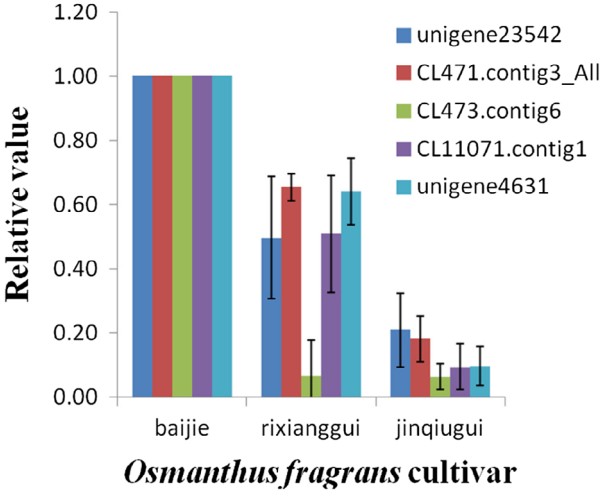


Figure 5. Q-PCR validation of differential gene expression of *Osmanthus fragrans*.

RNA-Seq technology was utilized to demonstrate the terpenoid biosynthesis of *O. fragrans* oil at a molecular level in this study. The transcriptome of *O. fragrans* was established, and differences in the expression of genes encoding the key enzymes in terpenoid synthesis between different types of *O. fragrans* were analyzed using RNA-Seq. All of the known enzymes involved in ionone synthesis were verified from the library. DEGs of *Osmanthus,* including "jinqiugui," "baijie," and "rixianggui" identified as one ispH, two GPPS, and one ZDS were screened, and 16 of a total 171 CCD unigenes were recognized as potential candidate genes that could be responsible for ionone biosynthesis. These results have practical significance for studies of *Osmanthus* oil in the fields of genetic control, metabolism regulation, and *in vitro* synthesis, among others.

Using transcriptome analysis, the molecular mechanisms of changes in the β-ionone concentration of *Osmanthus fragrans* were preliminarily examined. This provides not only a basis for further scientific investigation of the molecular mechanism within *O. fragrans* oil biosynthesis, but also a method for the identification and analysis of the genes involved in secondary metabolism in other non-model plants.

3. Experimental

3.1. Sample collection and preparation

"Jinqiugui" cultivars of *Osmanthus fragrans* var. thunbergii, "baijie" cultivars of *Osmanthus fragrans* var. latifolius and "rixianggui" cultivars of *Osmanthus fragrans* var. semperflorens were obtained from Hangzhou City *Osmanthus* Base, Zhejiang Province, China in Oct 2015. "Jinqiugui," "baijie," and "rixianggui" were in the early flowering phase (approximately 30–50% in bloom). The materials were frozen immediately in liquid nitrogen. Then part of materials was kept at −80 °C for total RNA extractions, another was analyzed by GC–MS.

3.2. Analysis of GC–MS

The volatile compounds in "jinqiugui," "rixianggui," and "baijie" were analyzed by GC–MS using headspace solid phase microextraction (HSSPME). Freshly flowers were placed into headspace sampling unites. GC–MS analysis was performed with a Trace DSQ instrument (Thermo Finnigan, California, U.S.A.) equipped with a capillary column of TR-5 MS (0.25 μm, 30 m × 0.25 mm; Keruihai, Beijing, China). The column temperature was kept at 40 °C for 2 min, and temperature programming was 2 °C min^{-1} to 300 °C with a final hold of 5 min. Carrier gas, He (99.999%) at 1.0 ml min^{-1}; injector temperature, 250 °C; interface temperature, 250 °C; splitting ratio 10:1; EI, ion ionization energy, 70 eV; mass scan range, 50–450 amu; and 5 scans s^{-1}.

Compound was confirmed by the NIST98 MS library matching, and the relative content was detected by the normalization method of peak area by G170BA chemical workstation.

3.3. Total RNA extraction, cDNA library preparation and sequencing

Total RNA was isolated from three different individual plant flower buds in each variety using the pBiozol Total RNA Extraction Reagent (BioFlux, Tokyo, Japan) kit and then mixed by equal quantity. The purity and yield of each RNA sample were determined by

RIN, 28S/18S, and the absorbance at 260 and 280 nm with Agilent Technologies 2100 Bioanalyzer and Nanodrop ND-1000 spectrophotometer (NanoDrop, Wilmington, U.S.A.) (Supplementary file 7). cDNA libraries and sequencing were performed by staff at Beijing Genome Institute.

3.4. De novo assembly

The image data output from sequencing machine was transformed by base calling into sequence data, which was stored in fastq format. First, RNA-Seq *de novo* assembly was carried out with the short reads assembling program Trinity (http://trinityrnaseq.sourceforge.net/) with the following parameters:–seqtype fq–min_contig_length 100; –min_glue 3–group_pairs_distance 250;–path_reinforcement_distance 85–min_kmer_cov 3. For each library, short reads were first assembled into linear contigs by building K-mers ($K = 25$). The contigs that may have alternative splicing and other paralogous genes were clustered, combining the linear path with continuous nodes in a de Bruijn map, and thus the sequence of the unigene was produced. Redundant unigenes were then removed by TGICL (v2.1) with options -l 40 -c 10-v 20. Finally, two classes were formed: one is singletons with the prefix "unigene," and another is clusters with the prefix "CL." Following this prefix are the id numbers of each unigene.

3.5. Functional annotation

Blastx alignment (E-value < 10^{-5}) between unigenes and protein databases like NT, NR, Swiss-Prot, GO, and KEGG was performed. ESTScan will be introduced to decide its sequence direction, when a unigene happened to be unaligned to none of the above databases.

With NR annotation, we use Blast2GO program to get GO annotation of unigenes (v.2.5.0). After getting GO annotation for every unigene, we used WEGO software to do GO functional classification for all unigenes.

3.6. DEGs Analysis

The DEGs between the two samples were performed using the method of Audic S [25]. We used the ratio of Fragments Per Kb per Million (FPKM) fragments for the two samples at the same time. We choose those with FDR ≤ 0.001 and ratio larger than twofold change in expression for GO functional analysis and KEGG pathway analysis. After multiple testing corrections, the Q-value of ≤ 0.05 in a pathway was defined as a significant enrichment of the pathway in the DEGs.

3.7. Real-time PCR verification of sequencing data

Five candidate unigenes of CCDs were selected for verification using Real-time PCR by Step One™ Real-Time PCR System (Life technologies, Carlsbad, U.S.A.). Primers were synthesized by Invitrogen Biotechnology Co., LTD (Shanghai, China) (Supplementary file 8), and the expression level of 18 S rRNA was used as a reference.

Detection system was in a total of volume 10 μl and each tube contained 1.0 μl of diluted cDNA, 1.0 μl of reverse and forward primers, 2.8 μl of ddH_2O, and 5.0 μl of the PCR master-mixture (Thunderbird SYBR qPCR Mix, Toyobo, Osaka, Japan).The thermal cycling conditions were as following: 95 °C/1 min, followed by 40 cycles with 95 °C/15 s, 58 °C/20 s, 72 °C/20 s. A heat dissociation curve (72–95 °C) was checked to test the specificity of the PCR amplification following the final cycle of the PCR. The relative quantification analysis was performed by ΔΔCT method. "Baijie" was set to 1. All samples were run in triplicates in separate tubes and each cDNA sample was run in triplicate. All data were expressed as the mean SD after normalization.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Special Fund for Forest Scientific Research in the Public Welfare under [grant number 201404601]; Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD); Jiangsu Key Lab of Biomass-based Green Fuels and Chemicals under [grant number JSBGFC14014], and the 11th six talents Peak Project of Jiangsu Province under [grant number 2014-JY-011].

References

[1] G.Y. Wang, H. Zhou, and X.F. Zeng, *Food Sci Technol*. **5**, 12–14 (2006).
[2] H. Cao, Z.G. Li, and D.L. Shen, *Acta Hortic. Sin*. **36**, 391–398 (2009).
[3] W.Y. Kang, and J.M. Wang, *Nat. Prod. Res. Dev*. **22**, 807–811 (2010).
[4] F.C. Huang, P. Molna′r, and W. Schwab, *J. Exp. Bot*. **60**, 3011–3022 (2009).
[5] J.G. Lashbrooke, P.R. Young, S.J. Dockrall, K. Vasanth, and M.A. Vivier, *BMC Plant Biol*. **13**, 156–173 (2013).
[6] H.N. Mu, H.G. Li, L.G. Wang, X.L. Yang, T.Z. Sun, and C. Xu, *Genes Genomics* **36**, 777–788 (2014).
[7] L.M. Wang, M.T. Li, W.W. Jin, S. Li, S.Q. Zhang, and L.J. Yu, *Food Sci*. **114**, 233–236 (2009).
[8] C.P. Moehs, L. Tian, K.W. Osteryoung, and P.D. Della, *Plant Mol. Biol*. **45**, 281–293 (2001).
[9] B.K. Shin, J.H. Ahn, and J. Han, *Bioinorg. Chem. Appls*. **8**, 1–8 (2015).
[10] J. Zhao, Y. Liu, Q.Y. Li, and X.L. Zhang, *Chin J Biotech*. **29**, 41–55 (2013).
[11] G. Wang, and R.A. Dixon, *Proc. Natl. Acad. Sci. USA* **106**, 9914–9919 (2009).
[12] Y.Y. Hsiao, M.F. Jeng, W.C. Tsai, Y.C. Chuang, C.Y. Li, T.S. Wu, C.S. Kuoh, W.H. Chen, and H.H. Chen, *Plant J*. **55**, 719–733 (2008).
[13] T. Isaacson, G. Ronen, D. Zamir, and J. Hirsehberg, *Plant Cell* **14**, 333–342 (2002).
[14] M.J. Rodrigo, J.F. Marcos, and L. Zacarias, *J. Agric. Food Chem*. **52**, 6724–6731 (2004).
[15] C. Zhu, S. Yamamura, M. Nishihara, H. Koiwa, and G. Sandmann, *BBA-Gene Struct. Exp*. **1625**, 305–308 (2003).
[16] A. Rubio, J.L. Rambla, M. Santaella, M.D. Gomez, D. Orzaez, A. Granell, and L.G. Gomez, *J. Biol. Chem*. **283**, 24816–24825 (2008).
[17] F.C. Huang, G. Horvath, P. Molnar, E. Turcsi, J. Deli, J. Schrader, G. Sandmann, H. Schmidt, and W. Schwab, *Phytochemistry* **70**, 457–464 (2009).
[18] N. Dudareva, and E. Pichersky, *Biology of Floral Scent*. (Taylor and Francis Group, NewYork, 2006), p. 40.
[19] F. Schade, R.L. Legge, and J.E. Thompson, *Phytochemistry* **56**, 703–710 (2001).

[20] A. Dubois, S. Carrere, O. Raymond, B. Pouvrea, L. Cottret, A. Roccia, J.P. Onesto, S. Sakr, R. Atanassova, S. Baudino, F. Foucher, M.L. Bris, J. Gouzy, and M. Bendahmane, *BMC Genomics* **13**, 638–649 (2002).
[21] A.J. Simkin, S.H. Schwartz, M. Auldridge, M.G. Tarlor, and H.J. Klee, *Plant J*. **40**, 882–892 (2004).
[22] N. Dudareva, and E. Pichersky, *Plant Physiol*. **122**, 627–633 (2000).
[23] M.B. Pott, F. Hippauf, S. Saschenbrecker, F. Chen, J. Ross, I. Kiefer, A. Slusarenko, J.P. Noel, E. Pichersky, U. Effmert, and B. Piechulla, *Plant Physiol*. **135**, 1946–1955 (2004).
[24] N. Kolosova, N. Gorenstein, C.M. Kish, and N. Dudareva, *Plant Cell* **13**, 2333–2347 (2001).
[25] S. Audic, and J.M. Claverie, *Genome Res*. **7**, 986–995 (1997).